DOCUMENT RESUME

ED 313 495 UD 027 192

AUTHOR Fetterman, David M.; Haertel, Edward H.
TITLE A School-Based Evaluation Model for Accelerating the Education of Students At-Risk.
PUB DATE 89
NOTE 36p.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Educational Assessment; Elementary Secondary Education; *Evaluation Criteria; *Evaluation Methods; Evaluation Utilization; Formative Evaluation; *High Risk Students; Nontraditional Education; *Organizational Effectiveness; Program Development; Qualitative Research; Self Evaluation (Groups); Summative Evaluation
IDENTIFIERS *Accelerated Schools Movement

ABSTRACT

This paper presents ideas for the development and utilization of a comprehensive evaluation plan for an accelerated school. It contains information about the purposes of a comprehensive evaluation, the evaluation design, and the kinds of data that might be gathered and used. The first section, "An Approach to Evaluation: Multiple Purposes and Multiple Perspectives," includes the following topics: (1) "Formative and Summative Evaluation Purposes"; (2) "An Ethnographic Perspective"; (3) "Intracultural Diversity"; (4) "Contextualization"; (5) "Nonjudgmental Orientation"; (6) "Emic or Insider's Perception"; and (7) "A Quantitative Perspective." The second section, "Getting Started: The Baseline Description, Summative Design, and Formative Evaluation Procedures," covers the following topics: (1) "A Baseline Description of the Accelerated School"; (2) Developing an Overall Evaluation Plan"; (3) "Comparative Bases for Summative Evaluation"; (4) "Time"; (5) "Curriculum and Instructional Methods"; (6) "Decision Making"; and (7) "Establishing Formative Evaluation Procedures." The second section, "Monitoring Progress: Ongoing Evaluation of Student Learning," includes the following topics: (1) "Systematic Classroom Observations"; (2) "Portfolios of Student Work"; (3) "Homework"; (4) "Grades and Report Cards"; (5) "Standardized Tests"; and (6) "Summing Up." Four notes and 50 references are included. (JS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED313495

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☒ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
David Fetterman
Stanford Univ.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

A School-Based Evaluation Model for Accelerating the Education of Students At-Risk

David M. Fetterman
Edward H. Haertel
School of Education
Stanford University

UD 027 192

BEST COPY AVAILABLE

1/20/88

A School-Based Evaluation Model for Accelerating the Education of Students At-Risk

David M. Fetterman
Edward H. Haertel
School of Education
Stanford University

An evaluation plan should help an accelerated school to learn about itself, to remember what it has learned, and to act on that learning in a systematic way. It should provide for the evaluation of student achievement, but should also attend to school climate, effective organization and decision making, communication with parents and the community, and the success of all the specific, planned for changes that will flow from a school's commitment to bring students at-risk into the educational mainstream. Much of good evaluation is just common sense, but evaluation can also draw on a body of proven methods and models to help ensure its success.

In this chapter, we share some ideas about a comprehensive evaluation plan for an accelerated school. No such discussion can provide an exact prescription, because the evaluation model, like all aspects of the accelerated school, must be tailored to local needs and must capitalize on a given school's particular strengths. As Professor Henry Levin remarked at the November 1988 Conference on Accelerated Schools, "This isn't a 'cookie-cutter' approach." We will address the needs of a school just embarking on the path of acceleration, and discuss a few of the many approaches possible. Necessarily, not all the steps we describe will coincide with the experiences of existing accelerated schools.

In the first section of this chapter, we describe some of

UD 027 192

the short-term and long-term information needs a comprehensive evaluation plan can address, and then introduce the mix of qualitative and quantitative approaches that we recommend to address those needs. The second major section, on getting started, discusses baseline data and overall evaluation design, as well as the organization and procedures that a school needs to establish to implement a comprehensive evaluation. A third section, on monitoring progress, describes some of the sources of evidence that may be used to chart student progress and keep the acceleration effort on track, and the fourth section offers some brief concluding remarks.

An Approach to Evaluation: Multiple Purposes and Multiple Perspectives

The evaluation of an accelerated school project requires both qualitative and quantitative approaches, and must be multidimensional to capture the all-encompassing accelerated school approach. The goals of this approach are ambitious and long term; thus an accurate assessment of the school's achievements should be based on longitudinal data, with different kinds of information having prominence at various stages in the project's development. Ideally, the evaluation data should be formative in nature, feeding information back to the program personnel and participants throughout the life of the project. This approach not only helps to keep the program and the evaluation alive and responsive to change, but also makes it maximally useful to its most critical constituencies. A summative evaluation component is also necessary. It should be designed to meet the information needs of the program sponsors and other audiences.

Formative and Summative Evaluation Purposes

The term evaluation conjures up different images for different people -- in part because it refers to so many different kinds of activities. One key requirement for the evaluation model in this chapter is that it be participatory. All the teachers, staff members, administrators, and students are potential data providers, and all of them, even the older students, are potential data collectors. A large part of the evaluation involves merely organizing and using the collective intelligence of the people in the school to guide the process of school reform.

A second use of the term evaluation conjures up an evaluator who shows up at the school with a box of standardized tests, administers them to all the students, then goes off and writes a report. This image is a caricature, of course, but it does serve to introduce the distinction between formative and summative evaluation: The first, participatory kind of evaluation is formative; the second is summative. The world will be watching the accelerated school, and at some point, judgments will be made about how well it is working, with or without supporting data. The purpose of a strong summative evaluation design is to ensure solid, defensible data to document the school's record of accomplishment. Obviously, standardized test scores are not enough to do the job. We briefly describe some alternative data sources below.

Our caricature was also inaccurate in suggesting that the summative evaluation takes place at a single point in time: Someone from outside comes in, collects data, and leaves. A

sound summative evaluation design involves the school personnel and requires systematic, longitudinal observations.

In summary, formative evaluation is flexible, ongoing, and designed primarily for internal consumption. It is most useful during the initial stages of the acceleration reform, when events and policies are still in flux. It may involve systematic data collection--quick surveys of opinion, for example, or a brief questionnaire about the good and bad aspects of an in-service training session--but much of the data for the formative evaluation will come from informal conversations, observations, or group discussions. The purpose of formative evaluation is to change things for the better. For that reason, its findings may quickly become obsolete. Once problems are corrected, the data identifying those problems are no longer pertinent.

Summative evaluation generally relies more heavily on paper-and-pencil measures collected systematically according to a prespecified design. It provides information for external audiences as well as for internal use and provides a sort of bottom line -- not just about the school being evaluated, but also about accelerated schools in general.

An Ethnographic Perspective

Qualitative descriptions draw their substance from the data, but take their shape from specific ethnographic concepts. The ethnographic concepts most valuable to the study of students at-risk and dropouts include intracultural diversity, contextualization, nonjudgmental orientation, and an emic perspective. In practice, these concepts work together; but they are separated here for purposes of discussion. A brief description of each

concept will highlight the significance of using a qualitative approach in an evaluation of this nature.

Intracultural diversity. There are many kinds of students at-risk. Dropouts, for example, are not a homogeneous group. Research has identified a variety of basic classifications, including dropout, potential dropout, and push out (see Fetterman 1981b, 1987). Still other students (also dropouts) roam from one school to another because they "do not want to get in with the wrong crowd." Documenting and recording the diversity within the dropout population helps to delimit specific "treatments" -- the application of specifically tailored educational programs and practices to each subgroup or subculture. This approach helps educators meet the individual needs of each student -- including the needs of gifted and talented students, who often drop out of school because they find it boring. Only a handful of gifted minority dropouts surfaced during a three-year study of dropouts across the United States. Contrarily, a three-year study of gifted and talented children (Fetterman 1988a) discovered a large number of disenfranchised (but mainstream) dropouts. Although dropouts are disproportionately represented by specific groups, they come from various ethnic and socioeconomic groups.[1] An approach that recognizes and accommodates these differences can help to dispel simplistic stereotypes.

Contextualization. Contextualization is among the most valuable of the qualitative concepts used to appreciate and understand the dropout phenomenon. In fact, the knowledge gained by documenting the dropouts' environments allows the educator and evaluator to appreciate the intracultural diversity and individu-

al differences within the population.

In the three-year dropout study, providing the context for thought and behavior was instrumental to understanding why people acted the way they did. For example, a description of the urban (lower socioeconomic) dropout's physical environment gave policymakers some understanding about what schools had to compete with in the inner-city. A brief description about war-zones within schools and lucrative but illegal activities provided insight into why some students dropped out, why some pursued one path over another.

Contextualizing data can also help the researcher and policymaker to appreciate the significant but subtle gains that programs often make. During the dropout study, for example, the federal government wanted to close the doors of one program for dropouts because attendance was too low. Former dropouts were attending this experimental program at a 50% attendance rate. Policymakers compared this rate with the average daily attendance in the neighboring inner-city school and decided program attendance was too low. They were framing their decisions within an inappropriate context. These students were systematically different from students who attended the neighboring high school. The appropriate attendance rate with which to compare a former dropout's performance is zero. Given this comparison, an attendance rate of 50% is excellent. This type of contextual information contributed to an understanding of how much effort these former dropouts were putting into this opportunity. The same information was also useful for policy decision makers. In this case, it contributed to the decision to keep the program operat-

ing.

Nonjudgmental Orientation. A nonjudgmental orientation is requisite to describing the multifaceted context of at-risk learners. The researcher must put value judgments aside while listening to some students at-risk describe their disturbing family problems and neighborhood conditions. Their descriptions provide useful data with which to understand behavior and gauge improvements. Value judgments about family life-styles, local conditions, and cultural differences, however, do not contribute to the evaluation effort.

Evaluators are human beings with biases -- often very strong political and personal biases. However, they go to great lengths to make their biases explicit whenever possible. They also try to mix their human responses to very different life-styles and cultures with a dispassionate conceptual accuracy and attention to detail. Personal opinions should not distract the evaluator from the task at hand: capturing an honest portrait of the dynamics of human interaction. (See Fetterman 1986, 1989.)

Emic or Insider's Perception. The emic perception is the insider's or native view of reality. Qualitative evaluators spend a great deal of time with the people they study, learning their language or jargon, customs, and beliefs. Verbatim quotations provide a useful illustration of how the qualitative evaluator attempts to elicit the insider's view. A student who thinks he has been unfairly perceived as incompetent or irresponsible explains in his own words that he "is not dumb" and that he believes the teachers are "prejudice." Regardless of whether his view is an objective perception of reality, this student's per-

sonal perception helps us to understand the motivation for his behavior. People act on what they believe, and those actions have real consequences for their lives.

Qualitative evaluators strive to understand and record the insider's perception of reality. The evaluator is responsible for putting all the usually conflicting perceptions together to explain why and how communication, as well as miscommunication, occurs in a much larger framework. This cultural framework may include a school classroom, program, district, or educational demonstration project (Fetterman 1981b, 1988a, 1988b, 1989).

Just as a multitude of qualitative evaluation concepts and techniques contribute to an understanding of the dropout phenomenon, they can contribute to an understanding of an accelerated school. In addition to the concepts discussed above, a holistic orientation, symbols, structure, and function will be useful in the study of accelerated reform efforts. Many of these concepts and a host of other techniques are available to practitioners working to understand and evaluate programs for students at-risk. (For a more detailed discussion about the anthropological contributions to understanding disenfranchised students, see Erickson 1984, 1987; Fetterman 1982, 1987, 1988c, 1989; Gibson 1987; Hanna 1982, 1988; Jacob and Jordan 1987; Spindler 1974, 1988; and Tharp and Gallimore 1989.)

A Quantitative Perspective

One of the first tasks in establishing the accelerated school's evaluation function will be to decide on quantitative measures. We comment briefly on measures of school organization and climate, of attendance, and of learning outcomes, but other

categories might well be added -- for example, parental involvement in and satisfaction with the school.[2] All these areas are measurable using primarily objective, paper-and-pencil measures, but the quantitative evaluation could also employ systematic teacher interviews, classroom observations, and other information-gathering techniques. Of course, some of these data sources are commonly associated with qualitative evaluation approaches, as well.

The school climate and organization can be documented using interviews, climate surveys, and various kinds of archival data: What decision-making bodies have been established in the school? How often did they meet? Do the minutes indicate that they were effective? Do teachers find the school a rewarding place to work? On questionnaires, do teachers indicate that they are happy with their work? Given a checklist of possible school problems, how do they respond?

Students can also provide valuable information about classroom climate. Several inventories (e.g., Fraser, Anderson, & Walberg, 1982; Moos & Trickett, 1987) ask children to rate the degree of challenge and difficulty in their schoolwork, whether they know what they are supposed to be doing most of the time, whether their schoolwork is interesting, and so on. Brief interviews will provide additional data.

Measuring attendance is straightforward, and monitoring attendance for purposes of the evaluation should pose little additional burden on students or teachers. Along with attendance per se, the evaluation might attend to procedures for follow-up with parents when absenteeism is excessive. Qualitative and

quantitative approaches must complement one another. For example, it is important to attend not only to attendance policies that exist on paper, but also to how they are actually implemented.

For student outcomes, standardized achievement tests really are the coin of the realm, and there is general agreement that they can measure some -- though by no means all -- important learning outcomes (Stetz & Beck, 1981). These tests may be biased against at-risk learners, in part because the children may not understand the importance of the tests and thus may not try hard enough to show their best work. Continued vigilence is also essential to detect possible cultural bias in these tests. In any case, standardized tests will be a cornerstone of the summative evaluation. They might include tests like the Comprehensive Tests of Basic Skills (CTBS), Iowa Tests of Basic Skills (ITBS), Metropolitan Achievement Tests (MAT) or other achievement batteries. School profiles or achievement data from state testing and assessment programs (e.g., the California Assessment Program) might also be used.

Finding other kinds of quantitative student outcome measures may be difficult because measures that aren't published and widely used may not be viewed as trustworthy. At the very least, though, writing samples in different genres should be obtained (e.g., Ruth & Murphy, 1988). If time and resources allow, a few well-designed performance tests would be quite valuable as well (Stiggins, 1987). Summative student measures might also include some affective measures such as the Piers-Harris ("How I Feel About Myself"), and perhaps a survey of study skills and habits.

Getting Started: The Baseline Description, Summative Design, and Formative Evaluation Procedures

A Baseline Description of the Accelerated School

An ethnographic or comparable qualitative description of the program is requisite to establish a context in which to interpret student and program growth and development. It is also useful to document the kinds of complex changes that occur over time, from the start-up stages to the fully operational stage of development. (See Fetterman 1988a, 1988b, 1988c, and 1989 for a detailed discussion about the application of qualitative approaches to evaluation.)

The description of context should include the following kinds of baseline, implementation, and operational information: school location; community characteristics, including socioeconomics and parent involvement; school structure and organization; curriculum and instructional strategies; degree of participation by teachers and students in school decision making; school ethos or culture; student attitudes toward education and the school in particular; student academic self-concept and expectations for success; student achievement (grades, test scores); attendance, study habits, and values; and additional information.

Ideally, some of the same data might be collected at a neighboring school in a comparable setting, but without the accelerated school "treatment." These comparative data would provide a valuable, additional baseline for assessment of the accelerated school program.

Developing an Overall Evaluation Plan

Guided by basic qualitative concepts and by the formative and summative purposes the evaluation is to serve, the evaluator should select critical focal points for the evaluation. Each program must adapt its program and consequently its evaluation to the local setting. However, some commonplaces for any such program exist. These focal points include the summative evaluation design, time, standards, curriculum and instructional methods, and decision making.

Comparative Bases for Summative Evaluation. The summative evaluation will provide defensible data about a few key outcome measures. In order to show growth, the evaluation should obtain at least one set of baseline measurements. A series of baseline measures -- if available -- makes the design stronger. This approach leads to an interrupted time series design (Campbell & Stanley, 1973).

Another way to strengthen the summative evaluation design is by collecting data in another school that is more-or-less comparable to the accelerated school in most respects, but lacks an accelerated school program. Ideally, data would be collected in both the accelerated school and the control school near the beginning of the reform effort, and annually for several years thereafter. If the accelerated school's outcomes improve and the control school's outcomes remain stable, there are strong grounds for attributing the improvements to the enacted reforms. The best evaluation design in a given case will depend on the availability of baseline measures going back through time, as well as on the availability of one or more suitable schools willing to

serve as controls. If scores on some standardized test are available for cohorts of children going back a few years, the summative evaluation might use that same test to see whether after the reform the test scores exceed past trends.

The design for the summative evaluation will also reflect the specific goals the school commits to. It will say what indicators are going to show improvement and when that improvement should be visible.

Time. Programs have their own life cycle. Those programs that are just starting should not have to match the same standards that apply to mature and fully operational programs. Similarly, students vary in ability, maturity, and length of time in the program. Individual differences in pacing and learning rates should be accomodated within the boundaries of the accelerated program ideals. An assessment of a student's progress should be flexible enough to respond to both personal and programmatic timelines and expectations.

During the initial stages of the program, formative evaluation might focus on the implementation process, including the growth of the school community's commitment to change, emergence of consensus among staff on program goals, establishment of priorities and expectations, and enlistment of support from parents and other constituencies. Lively debate about the purposes, goals, and methods of the acceleration effort might be a healthy sign during this period. Later, perhaps during the first year or so following the inception of the project, the formative evaluation would focus on implementation, examining the instructional processes in place and the quality of student experiences.

Still later, in a summative phase, evaluation would turn to the assessment of student learning outcomes. Attention to student learning will be critical in both the formative and the summative evaluation, but focusing narrowly on student learning outcomes too early in the reform process could be a serious mistake.

In examining outcomes, especially for purposes of summative evaluation, specific dates or milestones should be established to measure success. The most obvious deadline is the end of elementary education. If one goal of this program is to close the gap between disenfranchised students and mainstream students before they enter secondary school, then this milestone is a minimum requirement. Yearly progress is another natural demarcation point in the students' program.

Formative evaluation of both processes and short-term outcomes must proceed on a more micro level. Teachers and counselors might have weekly or biweekly meetings in which they discuss the progress of each child in the program. This technique has been effective in programs for dropouts and potential dropouts and other students at-risk (Fetterman 1981), and would provide an efficient context in which to collect useful qualitative and quantitative data for the ongoing formative evaluation. As the accelerated school matures and innovations become routine, such meetings can also help to maintain vigilance against falling back into old ways of thinking and working, and can help to assure that high levels of effort and outcomes are sustained. Finally, an ongoing formative evaluation of achievement can help to minimize the chance that the summative evaluation will come as a surprise.

Standards. One of the most fundamental changes the accelerated schools program can make is to establish authentic, high expectations for student success. The expectations of teachers, counselors, parents, and even peers about a child's prospects for academic success may become internalized and may strongly influence the development of the child. To maximize learning, it is important to set and maintain realistic, high expectations for all learners and to indicate clearly the criteria for successful performance. For purposes of evaluation, the research design must include some method to make measurements over time in a comparable manner, so that change can be assessed. (It does not follow that evaluations must rely exclusively on commercial standardized achievement tests.)[3]

At the same time, too great an emphasis on performance standards can lead to pseudo-scientific quantification and the excesses of behavioral objectives. Without standards, programs and events tend to drift. But merely insisting on some kind of standard invites a sort of meaningless quantification: People will say "80% correct answers," for example, without first asking, How hard are the questions? Worse still, they may set standards for future performance without adequate information about present performance levels. If nothing else, the direction the program ought to be moving in is usually clear. For some formative evaluation purposes, just looking for improvement may be sufficient.

Curriculum and Instructional Methods. An ongoing evaluation of students' response to the curriculum and instructional approach is essential in this type of program. Ineffective or

unresponsive curricula will undermine the goals of the program by reducing students' motivation or providing a shaky foundation for their future development. Similarly, if a significant mismatch between student characteristics and teaching methods exists, the flow of information and ideas may stop. Classroom observation and student interviews are critical elements of any evaluation of this type of program: They place a check on maladaptive behavior and identify and reinforce adaptive strategies.

Much of the research on teaching suggests that fast-paced, teacher-centered, clear and organized direct instruction is most effective, at least for basic skills and lower-order learning outcomes (Brophy, 1986; Brophy & Good, 1986). This kind of instruction may be evaluated using a combination of classroom observations and student interviews sampling a representative range of ability levels. If adaptive or individualized curricula play a part in the accelerated schools intervention, a more complex design for assessing curriculum and instructional methods and outcomes at the individual level may be necessary.

Decision Making. Finally, a review of the degree of participation in the school's decision making is requisite to determine the effectiveness of an accelerated school program. This review is a goal in and of itself--a process goal. It will have implications for other programs and policies throughout the school. Participant observation techniques[4] are extremely useful in this case to document decision-making patterns, including changes in the school's power structure and human organization. Participatory decision making is different from teacher autonomy, although a degree of teacher autonomy is also important. Rather, shared

decision making should lead to coordination and consensus, reflected in such outcomes as school grading, attendance, and discipline policies that the school accepts, publicizes, and consistently enforces.

Decision making should extend beyond school curricular and policy decisions. Teachers should also play a role in decision making about the evaluation itself. (See Booth 1987 for a discussion about collaborative evaluation in a primary school.)

Establishing Formative Evaluation Procedures

The power of the accelerated schools approach comes from its concerted focus on content, process, and structure. For that reason, it is critical that all the many specific changes in curricula, instructional practices, school governance, parent involvement, and so forth accomplish their separate objectives and advance the common purpose of academic acceleration. Clear goals should exist not only at the level of the school as a whole, but also at the level of components within the school. Ideally, each new program or practice should be a response to some specific problem or need. After a need is identified, a plan of action should be developed to address that need, and part of that plan should lay out a formative evaluation, following procedures established early in the life of the acceleration project. The process of change is complex, and systematic monitoring of the intended and unintended consequences of changes can help to keep things moving in the right direction.

Are the new curricula, instructional procedures, and other arrangements functioning as intended? Is one innovation working at cross-purposes with another? Even quick, informal evaluation

activities require some planning and resources, and so their procedures should be sketched out at the time when the innovation is first put into place. What kind of evidence will be collected to judge whether a new curriculum is achieving its aims, whether shared governance is working, or whether a teacher in-service training session is worth the time and money it costs? Who will be responsible for seeing that these evaluations are carried through? Who will collect and record the data? When will be the right time to look back and take stock? Should criteria be established in advance against which success will be judged? No set answers exist, but these questions should have routine consideration as the school initiates new programs.

Records need to be kept of formative evaluation decisions, but the answers to questions about evaluation evidence, timing, procedures, and standards needn't be carved in stone. The evaluation must remain flexible, and alert to unintended and unanticipated outcomes, as well as to those planned for. If early findings suggest some revision, an entire formative evaluation plan might have to be scrapped and rewritten.

There must be an explicit allocation of human and material resources to support both formative and summative evaluation functions. If seed money for the acceleration effort can be obtained from government agencies or foundations, those sponsors will probably be willing to commit some resources to the evaluation, as well. With or without external support, responsibility for the evaluation of each component must be assigned to some individual within the school, typically a faculty member directly involved with that component. Conducting the evaluation should

be regarded as a nontrivial part of the faculty member's duties.

Ideally, an accelerated school's evaluation will be informed by expertise from outside the school, perhaps from a nearby college or university, the school district's department of research and evaluation, or a state-run educational services unit. External support may take the form of direct assistance on specific issues of evaluation design, procedures, or instrumentation; as well as in-service training to improve the faculty's own capacity. Sustained involvement of one or more evaluation experts from outside the school is not indispensable, but may improve both the quality and the credibility of the evaluation.

Monitoring Progress: Ongoing Evaluation of Student Learning

We turn next to some important examples of formative evaluation evidence concerning student outcomes. Once again, these examples are by no means exhaustive.

Systematic Classroom Observations. Teachers probably rely more on their own ongoing classroom observations than on any other source of information in deciding when to ask a question, who to call on, when to interrupt the lesson to stop a conversation, even where to stand in the room (Stiggins & Bridgeford, 1986). During seatwork, a teacher may walk from desk to desk, noting who has the work well in hand and who needs a little more help. By mid-October, teachers have a pretty good idea about which students are best in math, what kind of activities each most enjoys, and in what corner of the room mischief is most likely to start.

Teachers can use a few simple techniques to improve the reliability and usefulness of these classroom observations so

that they become more useful for grading and even for evaluating new curricula and instructional methods. One technique is to use some simple system for systematic recording of classroom observations. A quick note jotted down on an index card, together with the date and a child's name, can be filed at the end of the day or the end of the week. When the time comes for grading or parent conferences, the teacher can pull out and use these notes about a child's work. During classroom discussions, a sheet of paper with a grid of squares corresponding to the locations of students' desks can be used to tally participation.

These recording systems can also alert teachers to how well they observe the class. Some children will always need more attention than others, and insisting on uniform numbers of observations for all is unnecessary, but no children should be completely neglected. A third concern, in addition to recording and coverage, is to note the typical, not the exceptional. We tend not to notice things expected and routine--the unexpected sticks in our minds. Moreover, teachers delight in the excitement of seeing children reach beyond themselves, and they tend to remember these moments. But to be most useful, systematic observations should also attend to the routine, typical performance levels of each child. Improvement from September to June in the average performance level may count for more than those two brilliant responses on November 7 and again on March 18.

Finally, the teacher can supplement systematic observations of routine classroom activities with occasional performance tests. Children may or may not know that they are being tested, depending on the teacher's objectives. A standard task should be

set, requiring children to make a speech, write a paper, or do some project that can be evaluated according to specified criteria. Rick Stiggins (1987, 1988; Stiggins, Conklin, & Bridgeford, 1986) has worked in this area, and offers the example of a project in which children must draw a picture illustrating a food chain with at least four different organisms. Such assignments can also provide opportunities to integrate the different subject areas.

Portfolios of Student Work. Collecting systematic samples of children's work throughout the year is an old technique. One new twist, in connection with writing and literacy instruction, is to have children actively participate in selecting the work that goes into their portfolios, adding one item every two weeks, say. Each student might have a brief, individual conference with the teacher to review the portfolio several times a year. The student and teacher together could decide on a few individual goals for the student to work on. The teacher would keep the portfolios, of course, but the children might value the chance to look at their own work from weeks or months earlier and see how their skills have improved.

Homework. A considerable amount of research has addressed the value of homework (e.g., Barber, 1986; Paschal, Weinstein, & Walberg, 1984; Turvey, 1986). Interesting, varied homework assignments hard enough to challenge but not frustrate, collected and returned promptly with teacher comments, clearly seem valuable. Endless problem sets that even the teacher seems not to care much about clearly are not useful. In some situations, it might be well to have parents sign off on homework or contract to

spend some amount of time each night working on it with their children.

A teacher may find homework quite useful for diagnostic purposes, but beyond the question of whether or not it is completed, homework does not seem as useful for grading or summative evaluation. Children may need a chance to learn and practice without the constant threat of evaluation, and they should have help doing it. It is probably better to base grades on unassisted work, or perhaps, with cooperative learning approaches, on group participation.

Leinhardt and Greeno (1986) have found striking differences between beginning and expert teachers in Pittsburgh, Pennsylvania -- particularly in the routines they establish for checking mathematics homework. Expert teachers check homework following the same routine every day. Children may pass papers back one seat, correct papers as the teacher reads the key, then return them. Using a show of hands for the number of students missing each problem, the teacher can very quickly learn how well the class did overall, which problems caused the most difficulty, and which children need the most help. Novice teachers showed considerable variability from day to day in the methods they used to check homework. They generally spent a lot more time and ended up with less useful information than the experts.

Grades and Report Cards. Grades and report cards can be an extremely useful system for communication among parents, teachers, and students, as well as providing a performance record of a child's progress through school. To be most useful, though, grading should accord with an explicit, schoolwide policy. Most

experts oppose "grading on a curve." In fact, the idea of forcing a fixed grade distribution seems contrary to the whole spirit of acceleration. A given grade should represent some absolute standard of attainment, and if all children can meet the highest standard, they should all get A's (Haertel, 1987).

Some teachers might disagree that a grade should represent an absolute standard--that is just our opinion. At least four bases exist for assigning grades, and most teachers probably use some mixture of them. First is grading on a curve; second is grading on the absolute level of performance; third is grading according to effort; and fourth is grading according to improvement. All these approaches have value, which is why an explicit, schoolwide policy is essential to ensuring that everyone clearly understands just what grades mean in a particular school.

One possibility is more differentiated grading, with one grade for absolute performance and another for effort. A second possibility is to provide some other way to acknowledge and recognize significant effort on the report card, without confusing the message about the student's level of performance. Report cards should include written comments as well as letters and numbers, should indicate strengths as well as weaknesses, and where possible should suggest ways to improve. They should also be issued frequently. We recognize that all these suggestions could mean more paperwork for teachers, and we take that problem very seriously. But teachers already have to prepare report cards, and it would be well to make them maximally useful.

Standardized Tests. Our last category under formative evaluation of student outcomes is standardized tests. Far more

research has focused on these tests than on teacher-made tests, much of it fairly discouraging. A study several years ago by the Center for the Study of Evaluation at UCLA (Dorr-Bremme, Herman, & Doherty, 1983; Herman & Dorr-Bremme, 1983) found that when teachers were asked what they used standardized test scores for, they said that they generally received them too late in the year to make much use of them, but that the principal and central office personnel needed them. When the researchers asked principals the same question, the principals said that they didn't use the standardized test results for much of anything, but that the teachers needed them. Students and teachers attend much more to classroom tests (Haertel, Ferrara, Korpi, & Prescott, 1984).

The problem with using standardized tests is more than just the time lag. These tests do not tie closely enough to the particular instructional topics a teacher is working on at any given time. In order to ensure the widest possible market for their tests, test publishers tend to focus on knowledge and skills that nearly all districts teach. These core skills are undeniably important, and giving a standardized, external test every year or two seems a sensible way to chart progress. At the school level, these tests can serve to anchor the school's achievement levels and standards to the rest of the world. But standardized tests must be supplemented with measures of other important learning outcomes and must not have undue influence in curricular decision making (Haertel, 1989).

Summing Up

We have discussed the purposes of a comprehensive evaluation, the evaluation design, and the kinds of data that might be

gathered and used. We have also observed that an effective, ongoing formative evaluation of learning outcomes can help to assure that summative results do not come as a surprise. It remains to consider the reporting of evaluation findings as program milestones are reached. If all goes well, the final, bottom-line assessment should be an occasion for celebration. It may show some remaining deficiencies, but will almost certainly show significant movement in the right direction.

It may be useful to prepare two or more summative evaluation reports, targeted to different audiences. The school will derive most use from a highly detailed report, documenting specific areas of strength and weakness, perhaps even listing particular learning objectives that need more attention. The district, any sponsoring agencies, and other professionals starting or studying accelerated schools may need a report that presents specific, technical information, but at a greater level of generality, omitting many details. Finally, a briefer, nontechnical summary may best serve parents, the news media, and the community. Evaluation reporting and use is a highly researched area, and a trained evaluator can offer advice to help ensure that the final report does not collect dust on a shelf.

Conclusion

When Professor Levin asked us to write about this topic, we were committed in a general way to the idea of accelerated schools, and readily agreed. As we have read more in this area, and as we have worked and talked with people on the Accelerated Schools Project, we have become increasingly convinced that the philosophy and the approach are correct.

This is a challenging educational effort. It is also a much needed one if we are to meet the needs of students at risk and close the gap between disenfranchised and mainstream students. An effective, comprehensive design for formative and summative evaluation can help to assure successful implementation. However, the evaluation of this effort will have to be as novel and refreshing as the program itself, if is to be responsive to this challenge.

Notes

1. Ogbu (1987) presents an excellent discussion about minority intracultural diversity, focusing on immigrant and nonimmigrant minorities.

2. See Logsdon et al. (1988) for a discussion about problems evaluating parental involvement.

3. See Fetterman (1981a) for a discussion about problems with standardized tests in evaluations of dropout programs.

4. See Fetterman (1989) for a detailed discussion about ethnography, including participant observation and non-participant observation. See also Spradley and McCurdy (1972) and Pelto (1970). Maxwell et al. (1986) present a useful example of how to mix qualitative and quantitative approaches in a single study.

References

Barber, B. (1986). Homework does not belong on the agenda for educational reform. Educational Leadership, 43(8), 55-57.

Booth, E. O. (1987). Researcher as Participant: Collaborative Evaluation in a Primary School. In D. M. Fetterman (Ed.), Perennial Issues in Qualitative Research (special topic edition, Education and Urban Society), 20(1), 55-85.

Brophy, J. E. (1986). Research linking teacher behavior to student achievement: Potential implications for instruction of Chapter I students. In B. I. Williams, P. A. Richmond, & B. J. Mason (Eds.), Designs for compensatory education: Conference proceedings and papers (pp. IV-121 to IV-179). Washington, DC: Research and Evaluation Associates.

Brophy, J. E., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd ed., pp. 328-375). New York: Macmillan.

Campbell, D. T., & Stanley, J. C. (1973). Experimental and quasi-experimental designs for research. Chicago: Rand McNally. (Originally appeared in N. L. Gage (Ed.), Handbook of research on teaching (2nd ed.), 1963)

Dorr-Bremme, D. W., Herman, J. L., & Doherty, W. (1983). Achievement testing in American public schools: A national perspective. Los Angeles, CA: University of California, Center for the Study of Evaluation.

Erickson, F. (1984) . School literacy, reasoning, and civility:

An Anthropologist's Perspective. Review of Educational Research, 54(4):525-544.

Erickson, F. (1987). Transformation and School Success: The Politics and Culture of Educational Achievement. Anthropology and Education Quarterly, 18(4):335-356.

Fetterman, D.M. (1981a). Blaming the Victim: The Problem of Evaluation Design and Federal Involvement, and Reinforcing World Views in Education. Human Organization, 40(1):67-77.

Fetterman, D. M. (1981b). Study of the Career Intern Program: Final Report -- Task C: Program Dynamics: Structure, Function, and Interrelationships. Mountain View, CA: RMC Research Corporation.

Fetterman, D.M. (1982). Ethnography in Educational Research: The Dynamics of Diffusion. Educational Researcher, 11(3):17-29.

Fetterman, D.M. (1986). Conceptual Crossroads: Methods and Ethics in Ethnographic Evaluation. In D.D. Williams, (ed.) Naturalistic Evaluation. New Directions for Program Evaluation, no. 30. San Francisco, CA: Jossey-Bass, pp. 23-36.

Fetterman, D. M. (1987). Ethnographic Educational Evaluation. In G.D. Spindler (Ed.), Interpretive Ethnography of Education: At Home and Abroad (pp. 79-106). Hillsdale, NJ: Erlbaum.

Fetterman, D.M. (1988a). Excellence and Equality: A Qualitatively Different Perspective on Gifted and Talented Education. Albany, NY: State University of New York Press.

Fetterman, D.M. (1988b). Qualitative Approaches to Evaluating Education, Educational Researcher, 17(8):17-23.

Fetterman, D.M. (1988c). Qualitative Approaches to Evaluation in Education: The Silent Scientific Revolution. New York, NY: Praeger Publishers.

Fetterman, D.M. (1989). Ethnography: Step by Step. Newbury Park, CA: Sage Publications.

Fraser, B. J., Anderson, G. J., & Walberg, H. J. (1982). Assessment of Learning Environments: Manual for Learning Environment Inventory (LEI) and My Class Inventory (MCI) (3rd version). Perth: Western Australian Institute of Technology.

Gibson, M. (1987). The School Performance of Immigrant Minorities: A Comparative View. Anthropology and Education Quarterly, 18(4):262-275.

Haertel, E. (1986, April). Choosing and using classroom tests: Teachers' perspectives on assessment. Paper presented at the meeting of the American Educational Research Association, San Francisco.

Haertel, E. (1987). Scores and Scales for School Achievement. Studies in Educational Evaluation, 13, 61-71.

Haertel, E. (1989). Student achievement tests as tools of educational policy: Practices and consequences. In B.R. Gifford (ed.) Test Policy and Test Performance: Education, Language, and Culture. Boston: Kluwer Academic Publications.

Haertel, E., Ferrara, S., Korpi, M., & Prescott, B. (1984, April). Testing in secondary schools: Student perspectives. Paper presented at the meeting of the American Educational Research Association, New Orleans.

Hanna, J.L. (1982). Public Social Policy and the Children's World: Implications of Ethnographic Research for Desegregated Schooling. In G.D. Spindler, (ed.) Doing the Ethnography of Schooling: Educational Anthropology in Action. New York, NY: Holt, Rinehart, and Winston, pp. 317-355.

Hanna, J.L. (1988). Not by Courts or Schools Alone: Evaluation of School Desegregation. In Fetterman, D.M., (ed.) Qualitative Approaches to Evaluation in Education: The Silent Scientific Revolution. New York, NY: Praeger.

Herman, J. L, & Dorr-Bremme, D. W. (1983). Uses of testing in the schools: A national profile. In W. E. Hathaway (Ed.), New directions for testing and measurement. No. 19: Testing in the schools (pp. 7-17). San Francisco, CA: Jossey-Bass.

Jacob, E. and Jordan, C. Explaining the School Performance of Minority Students. Theme Issue, Anthropology and Education Quarterly, 18(4):259-391.

Leinhardt, G., & Greeno, J. G. (1986). The cognitive skill of teaching. Journal of Educational Psychology, 78, 75-95.

Logsdon, D., Taylor, N., & Blum, I. (1988). It was a Good Learning Experience: The Problems and Trials of Implementing and Evaluating a Parent Participation Program. In D. M. Fetterman

(Ed.), Qualitative Approaches to Evaluation in Education: The Silent Scientific Revolution. New York, NY: Praeger.

Maxwell, J.A., Bashook, P.G., & Sandlow, L.J. (1986). Combining Ethnographic and Experimental Methods in Educational Evaluation: A Case Study. In D. M. Fetterman & M. A. Pitman (Eds.), Educational Evaluation: Ethnography in Theory, Practice, and Politics (pp. 121-144). Newbury Park, CA: Sage Publications.

McDermott, R. (1987). Achieving School Failure: An Anthropological Approach to Illiteracy and Social Stratification. In G.D. Spindler, (ed.) Education and Cultural Process: Anthropological Approaches. Second Edition. Prospect Heights, IL: Waveland Press, Inc., pp. 173-209.

Moos, R. H., & Trickett, E. J. (1987). Classroom environment scale manual (2nd ed.). Palo Alto: Consulting Psychologists Press.

Ogbu, J. (1974). The Next Generation: An Ethnography of Education in an Urban Neighborhood. New York, NY: Academic Press.

Ogbu, J. (1978). Minority Education and Caste: The American System in Cross-cultural Perspective. New York, NY: Academic Press.

Ogbu, J. (1981). Origins of Human Competence: A Cultural-Ecological Perspective. Child Development, 52:413-429.

Ogbu, J. (1982). Cultural Discontinuities and Schooling. Anthropology and Education Quarterly, 13(4):290-307.

Ogbu, J. (1983). Minority Status and Schooling in Plural Societies. Comparative Education Review, 27(2):168-190.

Ogbu, J. (1987). Variability in Minority Responses to Schooling: Nonimmigrants vs. immigrants. In G.D. Spindler and L. Spindler (eds.), Interpretive Ethnography of Education: At Home and Abroad. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 255-278.

Paschal, R. A., Weinstein, T., & Walberg, H. J. (1984). The effects of homework on learning: A quantitative synthesis. Journal of Educational Research, 78, 97-104.

Pelto, P. (1970). Anthropological Research: The Structure of Inquiry. New York, NY: Harper and Row Publishers.

Ruth, L., & Murphy, S. (1988). Designing writing tasks for the assessment of writing. Norwood, NJ: Ablex.

Spindler, G.D. (1974). Why Have Minority Groups in North America Been Disadvantaged by Their Schools? In G.D. Spindler, (ed.) Education and Cultural Process: Toward an Anthropology of Education. New York, NY: Holt, Rinehart, and Winston, pp. 69-81.

Spindler, G.D. (1988). Proceedings of the "What Do Anthropologists Have to Say About Dropping Out" Conference. School of Education, Stanford University, Stanford, CA.

Spradley, J. P. and McCurdy, D.W. (1972). The Cultural Experience: Ethnography in Complex Society. Palo Alto, CA: SRA, Inc.

Stetz, F. P., & Beck, M. D. (1981). Attitudes toward standardized

tests: Students, teachers, and measurement specialists. Measurement in Education, 12, 1-11.

Stiggins, R. J., & Bridgeford, N. J. (1986). The ecology of classroom assessment. Journal of Educational Measurement, 22, 271-286.

Stiggins, R. J., Conklin, N. F., & Bridgeford, N. J. (1986). Classroom assessment: A key to effective education. Educational Measurement: Issues and Practice, 5(2), 5-17.

Stiggins, R. J. (1987). NCME Instructional Module on design and development of performance assessments. Educational Measurement: Issues and Practice, 6(3), 33-42.

Stiggins, R. J. (1988). Revitalizing classroom assessment: The highest instructional priority. Phi Delta Kappan, 69, 363-368.

Tharp, R. and Gallimore, R. (1989). Rousing Minds to Life: Teaching, Learning, and Schooling in Social Context. Cambridge, MA: Cambridge University Press.

Turvey, J. S. (1986). Homework--Its importance to student achievement. NASSP Bulletin, 70(487), 27-35.